

【 专论 】

人工智能辅助病理图像标注的标准化建设

赖梓漩^{1,2}, 杨怡帆^{1,2#}, 任禹珂¹, 霍桂桃¹, 李双星¹, 张 颀¹, 杨艳伟¹, 高苏涛¹, 张 勇¹, 陈旭林¹, 林志^{1*}, 屈 哲^{1*}

1. 中国食品药品检定研究院安全评价研究所 药品监管科学全国重点实验室, 北京市重点实验室, 细胞及基因治疗药物质量和非临床研究与评价北京市重点实验室, 北京 100176
2. 中国药科大学 多靶标天然药物全国重点实验室, 江苏 南京 210009

摘要: 病理图像标注作为人工智能 (AI) 模型训练关键环节, 其准确性、一致性和标准化程度直接影响 AI 模型性能与临床应用价值。当前国内外缺乏统一完善的标注标准, 导致标注数据差异大, 限制数据共享复用及 AI 辅助病理诊断技术推广。系统梳理国内外相关标准与规范, 分析病理图像标注的核心要素 (标注对象、工具、评价指标) 及其面临的挑战, 如主观性差异、数据异质性、隐私安全与高成本等。国际上, 国际医学数字成像和通信 (DICOM) 标准及相关项目为标注质量控制与数据复用提供了参考; 国内有国家药品监督管理局 (NMPA) 发布的相关技术审查指导原则与《YY/T 1833.3-2022》标准, 以及《人工智能 面向机器学习的数据标注规程》等国家标准、山西省地方标准、中国通信标准化协会行业标准, 同时专家共识也在推进专科数据集建设, 在医疗器械监管、数据标注流程与质量评估等方面形成了初步标准体系。未来应加强 AI 辅助标注工具研发、专科标注规则制定与跨机构数据共享, 构建开放、安全、协同的标注生态, 推动数字病理发展和精准医疗实现, 以充分发挥病理 AI 模型的临床价值。

关键词: 人工智能; 数字病理; 数据标注; 标准化; 质量控制

中图分类号: R285.5 **文献标志码:** A **文章编号:** 1674-6376(2026)04-1089-09

DOI: 10.7501/j.issn.1674-6376.2026.04.001

Standardization of artificial intelligence-assisted pathological image annotation

LAI Zixuan^{1,2}, YANG Yifan^{1,2}, REN Yuke¹, HUO Guitao¹, LI Shuangxing¹, ZHANG Di¹, YANG Yanwei¹, GAO Suta¹, ZHANG Yong¹, CHEN Xulin¹, LIN Zhi¹, QU Zhe¹

1. National Center for Safety Evaluation of Drugs, State Key Laboratory of Drug Regulatory Science, Beijing Key Laboratory, Laboratory of Quality Control and Nonclinical Research and Evaluation for Cellular and Gene Therapy Medicinal Products, National Institutes for Food and Drug Control, Beijing 100176, China
2. State Key Laboratory of Natural Medicines, China Pharmaceutical University, Nanjing 210009, China

Abstract: As a critical step in training artificial intelligence (AI) models, the accuracy, consistency, and standardization of pathological image annotation directly determine the performance of AI models and their clinical applicability. Currently, the lack of comprehensive and unified annotation standards globally leads to significant variability in annotated data, which hinders data sharing, reuse, and the widespread adoption of AI-assisted pathological diagnostics. This paper systematically reviews existing standards and guidelines worldwide, analyzing core elements of pathological image annotation—such as annotation targets, tools, and evaluation metrics—along with challenges including subjective variability, data heterogeneity, privacy and security concerns, and high costs. Internationally, the Digital Imaging and Communications in Medicine (DICOM) standard and related initiatives provide references for quality control in annotation and data reuse. Domestically, guidelines such as those issued by the National Medical Products Administration (NMPA)

收稿日期: 2025-10-31

基金项目: 药品监管科学全国重点实验室课题项目 (2025SKLDRS0364)

作者简介: 赖梓漩 (2001—), 女, 硕士研究生, 研究方向为药物临床前安全性评价。E-mail: laizx_2001@163.com

#共同第一作者: 杨怡帆 (2001—), 女, 硕士研究生, 研究方向为药物临床前安全性评价。E-mail: yangyifan093@163.com

*通信作者: 林志 (1977—), 女, 研究员, 研究方向为药物临床前安全性评价。E-mail: linzhi@nifdc.org.cn

屈 哲 (1982—), 女, 研究员, 研究方向为药物临床前安全性评价。E-mail: quzhe@nifdc.org.cn

for technical reviews, the standard “YY/T 1833.3-2022”, along with national standards like “Artificial Intelligence—Data Annotation Specification for Machine Learning”, regional standards from Shanxi Province, and industry standards from the China Communications Standards Association, have been established. Additionally, expert consensus is advancing the development of specialized datasets, forming a preliminary framework of standards in areas such as medical device regulation, data annotation workflows, and quality assessment. Moving forward, efforts should focus on enhancing the development of AI-assisted annotation tools, formulating specialized annotation rules, and promoting cross-institutional data sharing to build an open, secure, and collaborative annotation ecosystem. This will drive the progress of digital pathology and the realization of precision medicine, ultimately unlocking the full clinical potential of AI-based pathological models.

Key words: artificial intelligence; digital pathology; data annotation; standardization; quality control

病理学诊断在疾病确诊中具有核心地位，是临床决策的重要依据。在现代临床实践中，数字病理学起着至关重要的作用，并且越来越成为实验室环境中的一项技术要求。全玻片成像的出现使病理学家更容易管理数字玻片图像并共享它们以供临床和非临床实验使用。与此同时，机器学习的空前进步实现了人工智能(AI)和数字病理学的协同作用，这提供了过去仅限于放射学和心脏病学的基于图像的诊断可能性^[1]。

在数字病理学领域，AI技术的飞速发展使病理图像分析带来了革命性的变化。AI模型在病理学领域的发展历程呈现出显著的技术演进轨迹，深刻改变了医学诊断格局。这一演进过程从专家系统开始，历经传统机器学习(ML)，最终发展到深度学习(DL)，体现了对提升诊断工具准确性与效率的持续探索。

AI辅助病理图像分析系统在疾病诊断、预后评估和治疗决策等方面展现出巨大的潜力，能够提高诊断效率、减少人为误差，并为精准医疗提供有力支持。病理图像标注作为AI模型训练的关键环节^[2]，其标注的准确性、一致性和标准化程度直接影响到AI模型的性能和临床应用价值^[3-4]。目前国内外对于病理图像标注缺乏统一、完善的标准，导致不同研究机构和企业生成的标注数据在质量、格式和标注规则等方面存在显著差异^[5]。标注人员的经验和主观性也会影响标注质量，导致标注结果不一致^[6]。这些差异不仅限制了标注数据的共享和复用，也阻碍了AI辅助病理诊断技术的广泛推广和临床转化。因此，制定一套科学、合理且具有广泛适用性的病理图像标注标准至关重要。

1 AI辅助病理图像标注的核心要素与标准化维度

1.1 标注对象与任务

一般来说，病理图像标注可分为3类，即切片级标注、图像级标注、像素级标注。切片级标注仅标

注切片整体病理属性(如“存在肿瘤”“无明确病变”)，操作简便，适用于AI分类任务。图像级标注标注特定区域的病理类型，用于AI目标检测任务(如病灶定位)，需将病理术语转化为可量化的AI特征标签(核异型性评分、腺体结构紊乱度等)^[7]，而像素级标注需病理专家准确勾勒出癌症区域边界，适用于AI语义分割任务，需满足像素级标注精度要求，且标注粒度需匹配模型卷积核尺寸，常用于病理图像分割^[8-9]。不同标注类型的选择需同步兼顾病理诊断金标准与AI模型训练需求，避免标注信息冗余或缺失。

标注不仅是图像勾勒，而且是将病理学诊断术语转化为AI可量化的特征标签。以肝淋巴瘤的辅助诊断为例，标注标准需将病理判读逻辑深度转化为数学特征。例如，病理学中诊断淋巴瘤的关键依据之一是中等至偏大的淋巴母细胞，细胞质稀少到中度、嗜碱性、并可能空泡化；核呈圆形、卵圆形、不规则形或卷曲，染色质有细微斑点。在标注时，这被映射为AI对细胞核几何特征及纹理特征的低方差量化。

1.2 标注工具

病理图像标注工具作为标准化标注规则落地实施的核心载体，核心聚焦于工具对标注规则一致性贯彻、全流程质量控制及完整追溯链条的系统性支撑能力。病理图像标注工具的选择应考虑其易用性、功能性、可扩展性以及与其他数字病理学系统的兼容性^[10]。一套契合标准化要求的病理图像标注工具，应具备以下5大核心技术特性。

其一，具备多尺度与多类型标注适配能力，能够覆盖切片级、区域级至像素级或实例级等不同维度的标注需求，实现对病理图像中各类目标的精准标注。其二，构建结构化标签体系与本体管理模块，支持层级化标签体系构建、医学术语标准化映射及标签版本的全生命周期管理，确保标注语言的统一

性与规范性。其三，集成全流程质控与审计追溯机制，涵盖操作行为的全要素日志记录、支持双人复核与仲裁的标准化工作流程，以及周期性抽样核查模块，为标注质量提供全流程保障。其四，具备优良的互操作性与标准化数据导出能力，能够实现与 DICOM 标准及 WSI 元数据的有效关联，并可灵活输出 JSON、XML 等结构化数据格式及掩膜图像等可视化数据格式，保障数据的跨平台流通与复用。其五，严格遵循安全与权限管理规范，实现基于角色的访问控制（RBAC）、敏感数据脱敏及数据传输加密等核心功能，确保标注数据的安全性与合规性。

在实际应用场景中，需结合具体需求场景科学遴选开源或商用工具。其中，开源工具（如 QuPath）

凭借其高度可定制性优势，更适用于科研探索场景及标注规则的快速迭代优化；而商用平台（如 Aperio ImageScope、HALO AI 等）则在多中心协同标注、高通量图像分析任务处理及审计追踪体系完善性等方面具备显著优势，更适配临床多中心研究及规模化诊断辅助场景。质量控制工具（HistoQC）能够自动化为病理图像质控，检测染色、聚焦、组织折叠等问题，常作为标注前处理步骤，提升数据标准化水平。

建立工具功能与标准化要求的深度适配机制，方能保障不同机构、不同标注人员产生的标注数据具备良好的一致性与可比性，为后续病理图像分析模型的训练及临床应用奠定可靠的数据基础。常用病理图像标注工具的核心功能对比见表 1。

表 1 常用标注工具功能及使用场景
Table 1 Functions and use scenarios of commonly used annotation tools

工具名称	主要功能	标注任务	类型	标准化支持特性	适用场景
QuPath	开源数字病理图像分析软件，支持 WSI，提供细胞计数、组织分割、批处理脚本扩展	分类、分割、检测、量化	开源	支持多种图像格式（包括 DICOM），可导出 GeoJSON/XML；支持多人协作与脚本化质控	研究场景常用，适用于算法开发与标注流程定制
Aperio ImageScope	商用数字病理查看与标注工具，提供注释、区域标记、测量工具	分类、区域标注、简单分割	商用	支持 SVS 等 Aperio 格式，标注可保存为 XML；集成于病理科常用系统	临床环境中常用，适用于诊断标注与教学
HALO AI	AI 数字病理分析平台，支持免疫组织化学、免疫荧光、组织微阵列等多模态图像定量分析	量化、分类、空间分析、细胞计数	商用	提供标准化分析模块，组织分类器、空间分析算法，支持质控流程与审计追踪	高通量研究、临床试验、毒理病理学
VisioPharm	数字病理图像进行细胞/区域的空间定量分析	细胞计数、细胞识别与分类、组织分割	商用	IHC 检测质量保证和标准化	临床诊断支持
HistoQC	自动化病理图像质控，检测染色、聚焦、组织折叠等问题	图像质控（预处理）	开源	可与标注流程集成，提升输入数据一致性	常作为标注前处理步骤，提升数据标准化水平

1.3 评价指标

评估 AI 模型性能的标准指标包括准确率、精确率（Pr）、召回率（Re）、F1 值以及曲线下面积（AUC）等^[11]。这些指标可以量化模型在诊断、分类和预测方面的能力。同时，鲁棒性也是一个重要的评价标准，确保模型在不同数据集和环境下的稳定性和泛化能力。Pr 指模型预测为阳性的所有样本中阳性样本的比例，也称为阳性预测值。Re，也称为真阳性率，指模型正确预测为阳性样本的样本比例，反映了算法的灵敏度。F1 评分是一个有价值的

评估指标，因为它平衡了准确度和回忆能力。对于二分类 AI 模型，常采用受试者工作曲线（ROC）及其 AUC 进行评估。AUC 量化了模型在所有可能分类阈值下的整体性能，其值越接近 1，表明区分能力越强。计算评估指标的方程如下。

$$Pr = TP / (TP + FP)$$

$$Re = TP / (TP + FN)$$

$$F1 = (2 \times Pr \times Re) / (Pr + Re)$$

TP 表示真阳性，FP 表示假阳性，TN 表示真阴性，FN 表示假阴性，表示模型预测阳性类为正（即 TP）或阴性（即 FN），

而阴性类分别预测为正（即 FP）或负性（即 TN）

1.4 核心挑战

病理图像标注面临诸多挑战，是标准化需要着力解决的问题。如高度依赖病理专家专业知识导致的主观性差异，使得不同专家对复杂或交界性病变的判断可能出现观察者间或观察者内差异；同时，像素级标注过程极其耗时耗力，精细标注单个全切片数字图像（WSI）常需要数小时。AI 模型通常需要大规模高质量标注数据，但获取面临困难，部分源于某些病理特征（如肿瘤浸润前沿或特定免疫细胞亚型）定义边界的模糊性。数据本身的异质性也是障碍，不同中心、扫描仪及染色批次导致的图像颜色、对比度和质量差异显著。此外，标注过程及数据存储、传输还涉及高度敏感的患者隐私信息，必须严格遵守隐私法规，这增加了复杂性。高昂的专家标注时间和专用工具费用构成了显著的成本负担。这些因素共同构成了病理图像标注工作的主要难点。

1.5 病理诊断逻辑与 AI 训练需求的衔接

病理诊断通常遵循“低倍定位→中倍判别结论→高倍确认细节（细胞学异型、核分裂象等）”的层级化判读逻辑。与之对应，AI 训练往往采用基于 WSI 的多尺度策略（如 WSI→ROI→patch），因此标注标准应明确：不同任务的最小标注单元（MAU）与模型训练的 patch 尺寸/分辨率（如 $5\times/10\times/20\times/40\times$ ）之间的匹配关系^[11]。实践上建议在标注说明文档中固化 3 类约束。

在病理图像分析中，通过将“诊断判读层级”转化为明确的“训练数据尺度与粒度规范”，可有效约束模型学习过程，从而显著降低跨机构数据异质性对模型泛化性能的影响^[12-13]。具体而言，该规范包括：空间尺度约束，即根据形态学特征确定最适观察倍数，例如评估“均质片状分布”的组织结构和“其他部位肿瘤性淋巴细胞聚集”的整体模式在中倍视野下更为稳定，而判断“细微斑点的核染色质”“中央核仁”的清晰度以及“高核质比”等细胞核特征，则需依赖高倍视野以保证可靠性；边界精度约束，即在语义分割任务中明确边界容许误差，例如按像素或微米计设定容差带^[11,14]；以及抽样策略约束，例如分类训练可采用弱标注（切片级或 ROI 级）配合代表性区域抽样，而分割训练则需基于“强标注（像素级）”进行^[15-16]。通过系统引入这些约束，能够使模型训练与临床诊断的实际层次和精度要求对齐，提升其在不同来源数据上的适应性与稳定性^[17]。

在制定相关标准时，需将标注容错率、质控策略与模型鲁棒性进行联动规定^[18]。由于临床诊断本身允许一定的观察者差异，而 AI 模型训练却对标签噪声高度敏感，标准中应明确以下方面：首先，定义一致性量化指标，如专家间一致性（Kappa/ICC）及分割任务中的 Dice 一致性^[19]；其次，建立系统的分歧处理机制，包括双人标注加仲裁、分层复核（由初级至高级）以及疑难病例触发专家会议；再者，阐明容错与模型鲁棒性的关联，针对边界模糊区域（如浸润前沿、炎症反应带等）可引入“容差带”“不确定类别”或“软标签”等标注策略，并在模型训练阶段采用噪声鲁棒损失函数或不确定性建模方法以提升鲁棒性。此外，标准应规范“AI 预标注-人工校正”协同流程的介入时机：例如在像素级分割任务中，可先由模型生成粗分割结果，再由专家进行精细修正，这能在保证质量的同时显著降低成本，但必须配套记录人工修改痕迹与版本追溯信息，以满足可审计性与可再现性的要求^[20]。具体标注流程见图 1。

2 国际 AI 辅助病理图像标注相关标准与规范

在国际范围内，AI 辅助病理图像标注的标准化建设主要围绕医学图像数据的互操作性、标注流程的规范性以及标注质量的可控性展开。尽管尚未形成专门针对病理图像标注的统一国际标准，但相关医学影像标准、AI 数据标注规范及研究实践，为病理 AI 标注的标准化提供了重要参考框架。

2.1 医学影像数据格式与互操作标准：国际医学数字成像和通信（DICOM）

在国际上，医学图像数据标注主要遵循国际标准化组织 ISO12052 的标准——DICOM 标准^[21]，DICOM 是全球公认的通信和管理各种医学影像和相关信息的标准。随着 WSI 技术在病理学中的广泛应用，DICOM 标准通过补充病理相关对象与元数据结构，逐步支持超大尺寸病理图像及其关联信息的规范化表达。

在病理图像标注方面，DICOM 标准通过对图像元数据、空间坐标体系及结构化注释的定义，为标注结果与原始图像之间建立了稳定的映射关系。这一特性为跨平台、跨机构的数据共享与复用提供了技术基础，也为将病理专家的定性描述转化为机器可读取的信息奠定了底层支撑。尽管 DICOM 本身并不规定具体的病理标注规则，但其在数据结构和互操作性层面的规范，对病理 AI 标注标准化具有基础性意义。

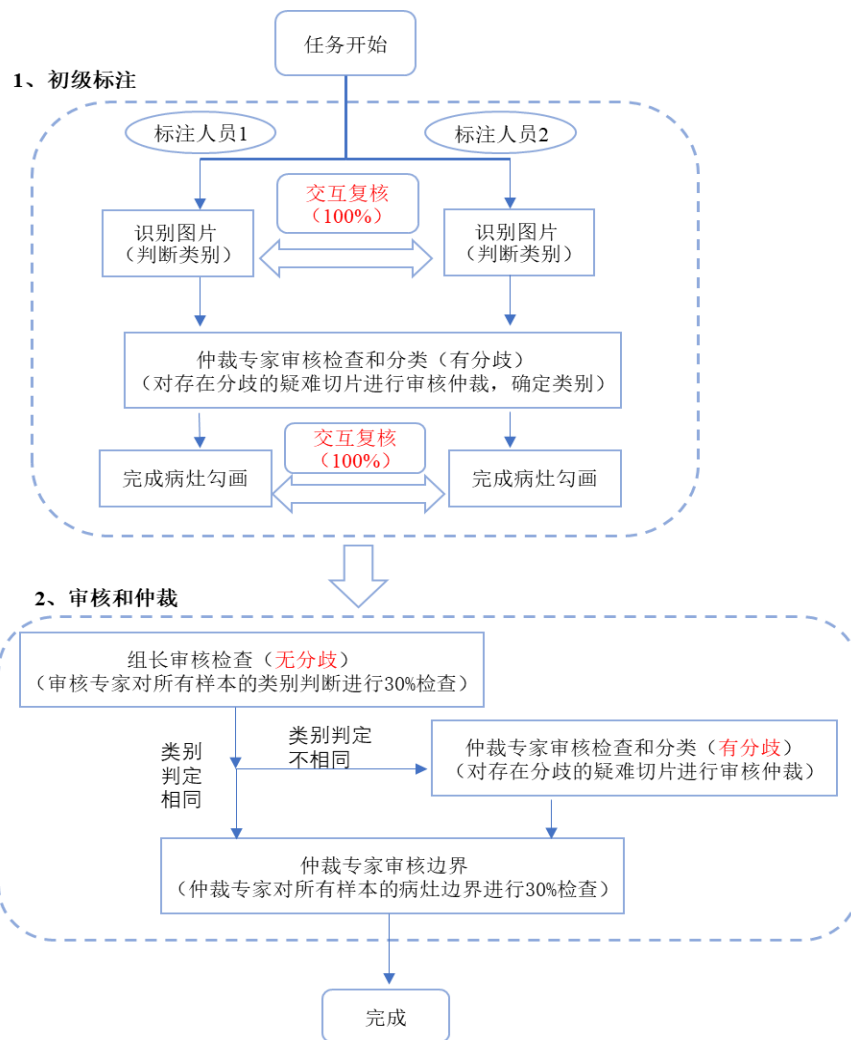


图 1 毒性病理图像数据标注的流程图

Fig. 1 Flowchart of toxicologic pathology image data annotation

2.2 面向 AI 的数据标注流程规范

中国信息通信研究院牵头的 AI 数据标注平台国际技术标准《Technical specification for artificial intelligence cloudplatform: Data annotation》^[22]在 ITU 成功立项, 该标准从 AI 数据标注全流程的角度出发, 共分为数据接入、数据标注、智能标注、数据交付和生产管理 5 个部分。由于数据标注在改进 AI 算法模型训练中的作用, 其重要性正日益受到关注, 结项标准围绕数据标注、云平台建设、模型测评等多个大模型产业化环节展开, 聚焦智能体感知评估、机器人流程自动化、代码生成、驾驶员行为检测等应用领域, 标志着中国在大模型国际标准领域取得了突破性进展, 推动了 AI 及相关技术标准的实现, 为 AI 辅助病理学图像数据标注提供参考。

2.3 标注质量控制与实践经验

数据标注是构建用于辅助检测和诊断的 AI 模型的重要环节。AI 模型开发需要高质量的学习材料, 尤其是在医学领域。Montezuma 等^[2]病理学家提供了一个在数字病理学中应用 AI 的实践指南, 强调了数据标注的重要性, 并讨论了数据标注过程中可能遇到的挑战。在标注质量控制方面, 建议采用多专家交叉验证模式, 例如对复杂病例引入第 3 位资深病理学家作为仲裁者, 以降低主观判断误差。同时, 针对不同 AI 任务需匹配差异化标注策略: 语义分割需逐像素标注组织类别, 如区分肿瘤与正常组织边界, 而目标检测则聚焦特定对象的定位与计数。通过提供清晰的指导, 可以帮助病理学家更有效地进行数据标注, 从而提高 AI 模型在疾病诊断和治疗中的应用。

Stadler 等^[23]构建了一个包含详细标注病理图像的数据库名为 DROID,用于训练用于肿瘤诊断的 AI 系统。DROID 项目通过覆盖检测、分割、分类等多任务标注,验证了“主动式”数据库构建的通用性,其经验表明标准化本体(如 SNOMED-CT)的应用可提升数据的跨场景复用能力。在标注质量控制方面,分层验证机制被证实有效。例如,DROID 项目中病理标注需经初级医师标注、高级医师确认,标注由住院医师与资深医师双重审核,此类流程可降低标注偏差,可为国内标准制定提供参考。

总体来看,国际上关于病理图像标注的标准化探索更多集中于数据结构、流程管理与质量控制方法层面,而非具体病理标注规则本身。这些经验为病理 AI 标注标准化提供了方法论基础,但在病理诊断逻辑与 AI 训练需求的深度融合方面,仍有待进一步系统化总结。

3 中国 AI 辅助病理图像标注相关标准与规范

目前我国尚未有专门针对毒性病理学图像数据标注的国家标准,但已逐步建立起以监管审评要求为导向、以国家/行业标准为技术支撑、以专科专家共识为补充的多层级标准体系,为 AI 辅助病理图像标注的规范化提供了制度基础。

3.1 监管导向下的病理 AI 标注质量要求

中国对 AI 医疗器械的监管日益规范,并发布了一系列相关标准和指南。国家药品监督管理局(NMPA)发布了多项针对 AI 医疗器械的技术审查指导原则,其中包括病理图像 AI 软件的性能评估和临床评价的关键审查要点^[24-26],在监管框架下,病理图像 AI 软件被普遍归类为 III 类高风险医疗器械,其安全性和有效性高度依赖训练与验证数据的标注质量。审评要点明确提出:训练和验证数据应覆盖不同医疗机构、扫描设备、染色批次及病理类型,以降低数据分布偏倚;标注过程应具备明确的操作规范、人员资质要求及完整的可追溯记录;在复杂或边界模糊的病理场景中,应通过多专家复核或仲裁机制构建相对稳定的“参考真值(reference standard)”。

上述要求将病理图像标注从单纯的数据准备环节,上升为直接影响医疗器械安全性和审评结论的关键技术环节,对标注流程的规范化与标准化提出了明确监管约束。

3.2 国家与行业标准对数据标注的体系化规范

NMPA 发布的国家标准《人工智能医疗器械质

量要求和评价第 3 部分:数据标注通用要求》(YY/T 1833.3-2022)^[27],这一标准适用于 AI 医疗器械在开发、训练和验证过程中所使用的标注活动。文件明确了标注任务的分类、标注规则、标注人员资质与分工、标注工具的功能与安全要求,以及标注环境对标注质量的影响,提出了标注结果准确性、一致性、精度、可理解性、可访问性、可移植性、保密性和可追溯性等质量特性,并详细规定了标注与质控流程,包括任务生成、分配、实施、质量控制和安全管理等环节,同时提供了对标注任务说明文档、标注与质控流程、标注工具及标注任务质量特性的评价方法,附录中还给出了多个具体标注任务示例和对 AI 辅助标注性能的评价指标,为 AI 医疗器械数据标注活动提供了全面的指导和规定。此标准从术语定义、任务文档、质量特性、流程、工具到评价方法,构建起完整规范体系,还给出多领域标注示例,对推动 AI 医疗器械发展意义重大,为构建适用于 AI 病理诊断的病理典型数据集提供参考。

国家标准《人工智能面向机器学习的数据标注规程》^[28]规定了面向机器学习的数据标注框架及流程。这个标准适用于指导 AI 领域面向机器学习的数据标注以及与之相关的研究、开发和应用等。它全面规范了以数据标注为主的实践与操作,适用于所有需要进行数据标注的机构、高校、企业、单位等。这个标准从初期计划、中期执行以及结果核查和输出 3 个阶段,分别梳理指导性框架、要求与步骤。标准要求标注过程遵循严格的规则,确保数据一致性、安全性和准确性,并通过多种验证方法(如机器验证、埋题抽查)保障质量。最终交付内容包括标注结果、规范文档及原始数据,同时需满足验收标准,未达标时需修正并重新质检。该标准适用于各类 AI 数据标注场景,为机器学习模型提供可靠的数据基础。

山西省地方标准《人工智能数据标注一般技术要求》^[29]规定了 AI 机器学习中数据标注通用技术要求的术语和定义、通用技术分类和通用技术要求。这个文件适用于 AI 机器学习中数据标注的企业、高校、研究机构、政府机关。其他行业和机构可以参照执行。

中国通信标准化协会发布的行业标准《面向互联网的医疗人工智能辅助决策 基于病理图像的辅助决策系统算法指标和测试方法》^[30]文件规定了面

向互联网的基于病理图像的医疗 AI 辅助决策系统的算法指标和测试方法,包括临床性能指标如 ROI 检测、图像分割、特征量化、病理分类分级等的具体指标及计算方法,以及其他性能指标如鲁棒性、泛化性、可再现性/可重复性,并详细描述了测试前的准备工作、测试流程包括临床性能测试和其他性能测试,以及测试后的封样流程,旨在通过全面系统的测试确保辅助决策系统的安全性和有效性,为监管机构提供有力的监管依据,推动该类产品的持续健康发展。其中数据标注应当考虑标注资源管理、标注过程质控、标注质量评估等要求。标注资源管理包括人员管理和基础设施管理。人员管理应当明确标注人员和仲裁人员的选拔、培训、考核等要求。基础设施管理应当明确标注场所、标注软件等要求。标注过程质控应当建立数据标注操作规范,明确标注人员、标注流程、临床诊疗规范(如临床指南、专家共识)、分歧处理(如仲裁人员、仲裁方式)、可追溯性(如数据、操作)等要求。标注质量评估应当明确人员、方法、指标、通过准则等要求。数据经标注后形成标注数据库,其样本类型可分为数据序列、单一数据(由多个数据块组成)、数据块(图像区域、数据片段)。样本量、样本分布等要求及风险考量与基础数据库相同。

3.3 专科化专家共识对病理标注规则的补充作用

除监管文件和技术标准外,近年来国内在病理 AI 领域逐步形成了若干专科化专家共识,围绕胃黏膜活检、实体瘤病理数据集建设等场景,对病理图像采集、标注规则及质量控制提出了更贴近临床实践的指导意见。该共识从胃黏膜活检和手术标本组织的病理文字图片数据采集及其质量控制、标注规则、标准流程等方面进行规定,为此类产业的数字集制造责任方提供思路。同时,专家们也在实体瘤病理数据集建设和数据标注质量控制方面达成共识,致力于推进实体肿瘤病理人工智能标准化数据建设^[31]。总的来说,这些研究都在探索降低人工注释工作量,提高学习效率的方法,有助于提升病理图像分析的准确性和效率。

4 结语

AI 辅助病理图像标注是连接病理诊断知识与 AI 模型训练之间的关键枢纽,其标准化建设对于提升模型性能、促进数据共享以及推动临床转化具有基础性意义。本文梳理了国内外相关标准与规范,从标注对象、工具、流程与质量控制等维度总结了

当前病理图像标注面临的主要挑战,并分析了国际经验与我国现有标准体系的特点与不足。

综合分析表明,病理图像标注标准化不仅是技术问题,更是病理诊断逻辑与 AI 训练需求协同设计的问题。单纯沿用通用数据标注规范,难以充分反映病理形态学的复杂性;而仅依赖专家经验,又难以满足大规模、高一致性数据集构建的需求。因此,未来标准化建设应重点围绕以下几个方面持续推进:建立面向不同病理亚专科的标注规则体系,将病理诊断术语系统映射为 AI 可量化、可解释的特征标签;明确不同 AI 任务(分类、分割、检测、量化)对应的标注粒度、精度与容错要求,实现标注规范与模型训练策略的协同;完善多专家参与的质量控制与仲裁机制,并在合规前提下推广“AI 预标注-人工校正”的半自动流程,以降低标注成本、提升一致性;加强跨机构数据共享与互认机制建设,在保障隐私与安全的前提下提升病理标注数据的复用价值。

通过构建科学、可操作、可审计的病理图像标注标准体系,有望进一步释放 AI 在数字病理和精准医疗中的应用潜力,推动病理 AI 模型从研究探索走向规范化、可持续的临床应用。

利益冲突 所有作者均声明不存在利益冲突

参考文献

- [1] Niazi M K K, Parwani A V, Gurcan M N. Digital pathology and artificial intelligence [J]. *Lancet Oncol*, 2019, 20(5): e253-e261.
- [2] Montezuma D, Oliveira S P, Neto P C, et al. Annotating for artificial intelligence applications in digital pathology: A practical guide for pathologists and researchers [J]. *Mod Pathol*, 2023, 36(4): 100086.
- [3] Feng S S, Zhang M C. A review of artificial intelligence in tumor pathology image analysis [J]. *Int J Comput Sci Inf Technol*, 2024, 3(1): 44-48.
- [4] Diao J A, Chen R J, Kvedar J C. Efficient cellular annotation of histopathology slides with real-time AI augmentation [J]. *NPJ Digit Med*, 2021, 4(1): 161.
- [5] Wang Y, Li N, Chen L M, et al. Guidelines, consensus statements, and standards for the use of artificial intelligence in medicine: Systematic review [J]. *J Med Internet Res*, 2023, 25: e46089.
- [6] Doğan R S, Yılmaz B. Histopathology image classification: Highlighting the gap between manual analysis and AI automation [J]. *Front Oncol*, 2024, 13:

- 1325271.
- [7] Ye J D, Luo Y H, Zhu C, et al. Breast cancer image classification on WSI with spatial correlations [A]. // 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) [C]. Brighton: IEEE, 2019.
- [8] Feng R W, Liu X C, Chen J T, et al. A deep learning approach for colonoscopy pathology WSI analysis: Accurate segmentation and classification [J]. IEEE J Biomed Health Inform, 2021, 25(10): 3700-3708.
- [9] Kadir M A, Tufiqur Alam H M, Maul P, et al. Modular deep active learning framework for image annotation: A technical report for the ophthlmo-AI project [J]. arXiv E Prints, 2024: arXiv: 2403.15143.
- [10] McGenity C, Clarke E L, Jennings C, et al. Artificial intelligence in digital pathology: A systematic review and Meta-analysis of diagnostic test accuracy [J]. NPJ Digit Med, 2024, 7(1): 114.
- [11] Gonzalez-Jimenez A, Lionetti S, Gottfrois P, et al. Robust T-Loss for medical image segmentation [J]. Med Image Anal, 2025, 105: 103735.
- [12] Lu M Y, Williamson D F K, Chen T Y, et al. Data-efficient and weakly supervised computational pathology on whole-slide images [J]. Nat Biomed Eng, 2021, 5(6): 555-570.
- [13] Shi J L, Zhang K L, Guo C Y, et al. A survey of label-noise deep learning for medical image analysis [J]. Med Image Anal, 2024, 95: 103166.
- [14] Yeung M, Yang G, Sala E, et al. Incorporating boundary uncertainty into loss functions for biomedical image segmentation [DB/OL]. (2021-10-31) [2025-07-16]. arXiv: 2111.00533. <https://arxiv.org/abs/2111.00533>.
- [15] Lu M Y, Chen T Y, Williamson D F K, et al. AI-based pathology predicts origins for cancers of unknown primary [J]. Nature, 2021, 594(7861): 106-110.
- [16] Campanella G, Hanna M G, Geneslaw L, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images [J]. Nat Med, 2019, 25(8): 1301-1309.
- [17] Tellez D, Balkenhol M, Otte-Holler I, et al. Whole-slide mitosis detection in H&E breast histology using PHH3 as a reference to train distilled stain-invariant convolutional networks [J]. IEEE Trans Med Imaging, 2018, 37(9): 2126-2136.
- [18] Dong J H, Zhang Y, Wang Q L, et al. Deep self-cleansing for medical image segmentation with noisy labels [J]. Med Phys, 2025, 52(10): e70007.
- [19] Zhang L, Tanno R, Xu M C, et al. Learning from multiple annotators for medical image segmentation [J]. Pattern Recognit, 2023, 138: 109400.
- [20] Khalili N, Spronck J, Ciompi F, et al. Uncertainty-guided annotation enhances segmentation with the human-in-the-loop [DB/OL]. arXiv:2404.07208. <https://arxiv.org/abs/2404.07208>.
- [21] National Electrical Manufacturers Association. The DICOM Standard: ISO 12052 [S/OL]. (2017-08) [2025-03-05]. <https://www.dicomstandard.org/current/>.
- [22] F. AICP-DA. Technical specification for artificial intelligence cloud platform: Data annotation (New): Initial draft [S/OL]. (2022-10-25) [2025-07-16]. <https://www.itu.int/md/T22-SG16-221017-TD-WP3-0051>
- [23] Stadler C B, Lindvall M, Lundström C, et al. Proactive construction of an annotated imaging database for artificial intelligence training [J]. J Digit Imaging, 2021, 34(1): 105-115.
- [24] Liu Y H, Yu W J, Dillon T. Regulatory responses and approval status of artificial intelligence medical devices with a focus on China [J]. NPJ Digit Med, 2024, 7(1): 255.
- [25] 国家药品监督管理局. 病理图像人工智能分析软件性能评价审评要点 [EB/OL]. (2023-07-10) [2025-07-16]. <https://ydcmdci.org.cn/article/263>. National Medical Products Administration. Review Points for Performance Evaluation of Artificial Intelligence Analysis Software for Pathological Images [EB/OL]. (2023-07-10) [2025-07-16]. <https://ydcmdci.org.cn/article/263>.
- [26] 国家药品监督管理局. 病理图像人工智能分析软件临床评价审评要点 [EB/OL]. (2023-07-10) [2025-07-16]. <https://www.ylzbzz.org.cn/index.php?m=content&c=index&a=show&catid=28&id=1580>. National Medical Products Administration. Review Points for Clinical Evaluation of Artificial Intelligence Analysis Software for Pathological Images [EB/OL]. (2023-07-10) [2025-07-16]. <https://www.ylzbzz.org.cn/index.php?m=content&c=index&a=show&catid=28&id=1580>.
- [27] 国家药品监督管理局. 人工智能医疗器械质量要求和评价 第 3 部分: 数据标注通用要求:YY/T 1833.3-2022 [S/OL]. (2022-08-17) [2025-07-16]. <https://std.samr.gov.cn/hb/search/stdHBDetailedid=E92817BCAF87FFFFE05397BE0A0A6860>. National Medical Products Administration. Artificial intelligence medical device-Quality requirements and evaluation-Part 3: General requirements for data annotation: YY/T 1833.3-2022 [S/OL]. (2022-08-17) [2025-07-16]. <https://std.samr.gov.cn/hb/search/stdHBDetailedid=E92817BCAF87FFFFE05397BE0A0A6860>.
- [28] 国家标准委. 人工智能 面向机器学习的数据标注规程: GB/T 42755-2023 [S/OL]. (2023-05-23) [2025-07-

- 16]. <https://std.samr.gov.cn/gb/search/gbDetailed?id=FC816D04FEB462EBE05397BE0A0AD5FA>. Standardization Administration of China. Artificial intelligence - Data annotation specifications for machine learning: GB/T 42755-2023 [S/OL]. (2023-05-23) [2025-07-16]. <https://std.samr.gov.cn/gb/search/gbDetailed?id=FC816D04FEB462EBE05397BE0A0AD5FA>.
- [29] 山西省市场监督管理局. 人工智能 数据标注一般技术要求: DB14/T 2464—2022 [S/OL]. (2022-06-14) [2025-07-16]. <https://std.samr.gov.cn/db/search/stdDBDetailed?id=E32E88480C25218FE05397BE0A0AC689>. Shanxi Provincial Administration for Market Regulation. Artificial intelligence-General technical requirements for data annotation: DB14/T 2464-2022 [S/OL]. (2022-06-14) [2025-07-16]. <https://std.samr.gov.cn/db/search/stdDBDetailed?id=E32E88480C25218FE05397BE0A0AC689>.
- [30] 中华人民共和国工业和信息化部. 面向互联网的医疗人工智能辅助决策 基于病理图像的辅助决策系统算法指标和测试方法: YD/T 6090—2024 [S/OL]. (2024-10-24) [2025-07-16]. <https://std.samr.gov.cn/hb/search/stdHBDetailedCNF?id=29962E7E0641A432E06397BE0A0A0BF0>. Ministry of Industry and Information Technology of the People's Republic of China. Medical Artificial Intelligence Auxiliary Decision-Making for The Internet-Algorithm Indicators and Test Methods of Auxiliary Decision-Making Systems Based on Pathological Images: YD/T 6090-2024 [S/OL]. (2024-10-24) [2025-07-16]. <https://std.samr.gov.cn/hb/search/stdHBDetailedCNF?id=29962E7E0641A432E06397BE0A0A0BF0>.
- [31] 于观贞, 陈颖. 实体瘤病理数据集建设和数据标注质量控制专家意见(2019) [J]. 第二军医大学学报, 2019, 40(5): 465-470.
Yu G Z, Chen Y. Establishment of pathological data set and quality control of labeling for solid tumor: expert opinion 2019 [J]. Acad J Second Mil Med Univ, 2019, 40(5): 465-470.
- [32] 滕伊洋, 孟娜, 侯敏博, 等. 数字组织图像分析和人工智能在毒性病理学中的应用进展 [J]. 药物评价研究, 2025, 48(4): 1050-1057.
Teng Y Y, Meng N, Hou M B, et al. Application advance of digital tissue image analysis and artificial intelligence in toxicologic pathology [J]. Drug Eval Res, 2025, 48(4): 1050-1057.

[责任编辑 刘东博]